Biology & Philosophy (2021) 36:3
https://doi.org/10.1007/s10539-020-09778-8

Is cancer a matter of luck?

Anya Plutynski[1]

Received: 7 January 2020 / Accepted: 29 December 2020 / Published online: 11 January 2021

Abstract

In 2015, Tomasetti and Vogelstein published a paper in *Science* containing the following provocative statement: "… only a third of the variation in cancer risk among tissues is attributable to environmental factors or inherited predispositions. The majority is due to "bad luck," that is, random mutations arising during DNA replication in normal, noncancerous stem cells." The paper—and perhaps especially this rather coy reference to "bad luck"—became a flash point for a series of letters and reviews, followed by replies and yet further counterpoints. In this paper, I critically assess Tomasetti and Vogelstein's argument, discuss the meaning of "luck" (or, better: "chance") in the context of the debate, and use this case study to address larger questions about methodological criteria for causal explanations of population level patterns in biomedicine.

Keywords Cancer · Explanation · Causation · Probability · Chance

Introduction

In 2015, Tomasetti and Vogelstein published a paper in *Science* containing the following provocative statement:

> … only a third of the variation in cancer risk among tissues is attributable to environmental factors or inherited predispositions. The majority is due to "bad luck," that is, random mutations arising during DNA replication in normal, noncancerous stem cells (2015, p. 78).

The paper—and perhaps especially this rather coy reference to "bad luck"—became a flash point for a series of letters and reviews contesting the claim (Various authors

Supplementary information The online version of this article (https://doi.org/10.1007/s10539-020-09778-8) contains supplementary material, which is available to authorized users.

✉ Anya Plutynski
aplutyns@wustl.edu

[1] Philosophy and Division of Biological and Biomedical Studies, Washington University in St. Louis, St. Louis, MO, USA

2015; Nunney and Muir 2015; Rozhok et al. 2015; Weinberg and Zaykin 2015; Wu et al. 2016), followed by replies from the authors, and yet further counterpoints (Tomasetti et al. 2017; Nowak and Waclaw 2017). Critics have argued that the authors made basic mistakes in logic and statistical reasoning, failed to consider additional endogenous causes of variation in average incidence, and ignored decades of public health research (Sornette and Favre 2015; Davy-Smith et al. 2016). As of this writing, there are currently several hundred papers in the scientific literature responding to Tomasetti and Vogelstein, in journals ranging from *Nature* to *Mutation Research—Genetic Toxicology and Environmental Mutagenesis*. Given the sheer number of critical responses to this paper already in the scientific literature, one might well ask whether this paper deserves any further attention. Was the authors' talk of "luck" in this context merely unlucky—an unusual lapse in judgment on the part of cancer researchers, not particularly deserving of philosophical attention?

On the one hand, it is surely true that Tomasetti and Vogelstein's paper has already received ample scrutiny. On the other hand, however, it seems worth making explicit how—if at all—this paper sheds any light on the matter of the senses in which variable cancer incidence is a matter of "chance." Thus, my aim here is to address the following questions, in turn: what exactly did the authors' talk of "luck" or "chance" in this context mean? Why did scientists critique the argument in such different ways, and which critiques hit the mark?[1] What lessons can we draw from this case about causal inference and explanation of population level patterns in biomedicine? Why are inferences regarding such matters so contested? Last but not least, what are the sources of "chance" (or contingency) in relative cancer incidence?

The "bad luck" debacle deserves philosophical scrutiny, not only in order to assess the strengths and weaknesses of Tomasetti and Vogelstein's particular argument, but also in service of addressing these broader questions. The aim of this paper is thus to address the above questions. In service of this aim, I first explain and describe Tomasetti and Vogelstein's claims and argument (section "The target of explanation: in what sense is cancer "bad luck"?"), and discuss a variety of scientific critiques of their reasoning and conclusions (section "Replies and controversies"). Second, I explain why—in their case—they were not warranted in drawing the causal inferences they did. This requires a brief philosophical discussion of causal inference in the context of population level outcomes, such as patterns of cancer incidence. I compare and contrast their argument with Doll and Hill's argument for the link between smoking and lung cancer, and Haldane's defense of "bean bag" genetics, and explain why by comparison Tomasetti and Vogelstein's reasoning falls short (section "Diagnosing the controversy: causation, chance, explanation and control"). Third, I explain that, even though their particular argument was flawed, there

[1] As we will see, there is some disagreement among the critics concerning the most serious flaw in Tomasetti and Vogelstein's argument. These differences at least in part overlap with disciplinary specialty, suggesting different pragmatic interests, methodological commitments, and perhaps also different views about causation, causal selection, and explanation, drove this debate (see also, e.g., O'Rourke et al. 2016; Brigandt 2013).

are, nonetheless, several ways in which one can meaningfully speak of cancer as being a matter of "chance" (section "On how cancer is a matter of chance").

Before I begin with a discussion of their argument, however, it is important to describe several distinct senses of "chance" at work in scientific discourse that different scientists appeal to in this debate. The phenomena cancer scientists typically seek to explain are not single events, but *types*, or averages over types of event or process (e.g., frequency of this or that mutation type in some class or subclass of cancers, or average lung cancer mortality in the UK, over a specific decade). One may speak of such types of event, or outcomes, as a matter of "chance" in a variety of ways.[2]

The first sense of "chance" is typically used in formal modeling of a given event or event type: e.g., one specifies that an event or event type is equiprobable with some alternative or set of alternatives. Second, such events or event types may be simply unpredictable; this is a "subjective" sense of "chance." One cannot as yet assign a probability or range of probabilities to such event types. Third, an event or event type may be relatively indiscriminate in its effects, for instance, on fitness or health outcomes. (For instance, errors in replication of DNA during somatic cell division are more or less indiscriminate in their effects.[3]) Fourth, chance is often used as a proxy for "probability." (For instance, with a higher number and rate of turnover of cells in some tissue, there is a greater probability (or "chance") of accumulation of mutations in such tissue, *ceteris paribus*).

This fourth sense of "chance" as merely "probabilistic" applies both to "endogenous" and "exogenous" (or extrinsic) factors, the latter of which Tomasetti and Vogelstein label "deterministic." Many exogenous factors exhibit a probabilistic relationship with cancer incidence; increased exposure to UV radiation in some population, for instance, increases the probability of cancer incidence in a population. So, *all* cancers are a product of "chance," if by this one means only that changing values of either endogenous or exogenous variables can change the probability of average incidence of this or that cancer in a given population. In this sense at least, whether "chance" is "more significant" than "deterministic" causes is an ill formed question. Arguably, this way of framing the matter is one major reason for confusion and dispute regarding Tomasetti and Vogelstein's argument, as we shall see, below.

A fifth sense of "chance" refers specifically to endogenous or "intrinsic" causal factors that increase relative vulnerability to disease. For instance, the proportion of infant deaths due to any and all intrinsic causes of susceptibility to congenital

[2] Unfortunately, when biologists talk of this or that class of event as being a matter of "chance," they are sometimes unclear about whether they intend to refer to our epistemic state (e.g., given our knowledge of various initial conditions, we can at best assign the events some probability), or a state of the world (i.e., the events are in fact not determined). Nonetheless, there are some contexts in which—at least with respect to relevant (known) macro-causal variables (e.g., environmental exposures, such as UV radiation)—biologists do have a very good estimate of the probability of various classes of outcome (e.g., melanoma). For the purposes of my discussion here, we can speak of these as "objective probabilities."

[3] Though there are some regions of the genome more prone to error than others, and some changes to the genome that tend to accelerate the onset of disease more quickly than others (Roberts and Gordenin 2014; Salk et al. 2010; Martincorena and Campbell 2015). For further discussion of the ways in which mutation is random, see, Merlin (2016).

disease shared by all human infants are sometimes referred to as those deaths that are "a matter of chance." Development is a complex process, and errors during this developmental process (whether errors in cell division, tissue formation, or what have you) can lead to congenital malformations, or to disease. Such congenital problems are often referred to as "a matter of chance," where the implicit contrast in such cases is with causal factors over which we have some control (potentially or actually)—such as environmental exposures, like smoking. Tomasetti and Vogelstein refer to the proportion of relative cancer incidence that (they claim) can be predicted in light of species-wide patterns of cell division across different tissues or organs as that proportion "due to chance." Errors in replication of DNA during cell division are referred to a matter of "chance" in this sense, in that they are a product of intrinsic features of cellular replication machinery, (and insofar as whether such errors yield cancer is to some extent a matter of probability. Most such mutations have no effect, but a few, cumulatively over time, do). That cancers arise in some tissues or organs on average more often than others is a matter of "chance" in this fifth sense, in that their relative occurrence is an intrinsic product of different rates of cell turnover (or, so Tomasetti and Vogelstein argue).

Tomasetti and Vogelstein are not always clear which of the several senses of "chance" above they have in mind, and sometimes they may refer to several at once. At different points, they seem to be referring to "chance" in the sense that mutations can be relatively indiscriminate in effect, as proxy for "probability," or as a matter of "intrinsic" or "baseline" relative risk, across tissue types. This last sense seems to be the core sense they are referring to, at least viz. the contrast with so-called "deterministic" causes. Unfortunately, the fact that they are not always precise led in part to criticisms of their views. Some of the major objections seem founded on false or confused understandings of their intended meaning.

Nonetheless, there is one major criticism that is absolutely spot on: namely, that partitioning the relative causal contribution of "chance" and "deterministic" factors to differential cancer risk in the manner they do is inadequately supported by the evidence they provide. Assigning anything like a specific value to the relative contribution of "luck" or "chance" (in Tomasetti and Vogelstein's sense) is problematic, in part because there the so-called "chance" and "deterministic" factors are not mutually exclusive, and in part because stem cell turnover is only one of many endogenous factors at work in relative cancer incidence across tissues and organs. In other words, stem cell turnover is not a stable and proportionate cause of relative cancer incidence, but highly contingent in its effects, in ways that vary across tissues and organs. Let us turn now to their positive argument.

The target of explanation: in what sense is cancer "bad luck"?

Before analyzing Tomasetti and Vogelstein's argument, it's important to first specify the target of explanation. Their interest was not in explaining each individual's chance of cancer, or average lifetime incidence of all cancers, but differences in relative incidence of different cancer types (more precisely, cancers arising in different tissues and organs—ovarian, breast, or prostate cancer, for instance). That is, they

were *not* arguing that each person has an equal probability of getting cancer. Their target was instead differences in average (lifetime) cancer incidence across cancer types. Average lifetime cancer risk across tissue types varies, in some cases, by orders of magnitude.

Tomasetti and Vogelstein took these significant differences to be puzzling, because while some of the differences in lifetime risk "are associated with well-known risk factors such as smoking, alcohol use, ultraviolet light, or human papilloma virus (HPV)" (2015, p. 78), the most significant differences in incidence do not track differential environmental exposures. For instance, "such exposures cannot explain why cancer risk in tissues within the alimentary tract can differ by as much as a factor of 24 [esophagus (0.51%), large intestine (4.82%), small intestine (0.20%), and stomach (0.86%)]…" (2015, p. 78) Nor can such exposures explain why "cancers of the small intestinal epithelium are three times less common than brain tumors … even though small intestinal epithelial cells are exposed to much higher levels of environmental mutagens than are cells within the brain, which are protected by the blood–brain barrier." (2015, p. 78) In other words, if environmental exposures were the major causes of such differences, (they claim), then we should expect, e.g., much higher rates of cancer in the intestinal epithelium than in the brain. But these differences in incidence are in fact the reverse of what one might expect if environmental factors are the main difference maker.

Tomasetti and Vogelstein's answer to this puzzle is as follows: differences in average incidence are (largely) due to "stochastic[4] effects associated with the lifetime number of stem cell divisions within each tissue." More precisely, differences in number and rates of stem cell division in different tissue types is largely responsible for the large magnitude in difference between lifetime cancer risk in rare versus common cancer types. They go yet further, however, when they claim the following:

> We show here that the stochastic effects of DNA replication can be numerically estimated and distinguished from external environmental factors. Moreover, we show that these stochastic influences are in fact the major contributors to cancer overall, often more important than either hereditary or external environmental factors. (2015, p. 78)

[4] I take Tomasetti and Vogelstein to be using the term "stochastic" in the sense that per base pair somatic mutations occur at a regular rate, sometimes called the "background" mutation rate. There are different causes of this outcome, or means by which perfect replication fails to come about. For instance, errors due to random polymerase misincorporation are estimated to be 7.6×10^{-10} ($\pm3.8\times10^{-11}$) per base per cell division (Tomasetti et al. 2013). As we will see, some readers interpret their use of the term "stochastic" as "random," in the sense that, given the same initial conditions, all causes of mutation, and all outcomes (e.g., base pair changes, inversions, deletion, chromosomal duplications, etc.) are equiprobable. This seems unlikely, however, as it's widely known that some causes or error, and some types of mutation, are more common than others. We can predict which elements of the genome are most subject to error, yielding mutations of some types more often than others (Kimsey et al. 2015; Kunkel 2009; Fromme and Verdine 2004; Collins 2005; Gold 2017). Biologists sometimes use the term "stochastic" to refer to processes whose outcomes can be modeled as a random sampling procedures, but Tomasetti and Vogelstein make no such claim (at least explicitly) here.

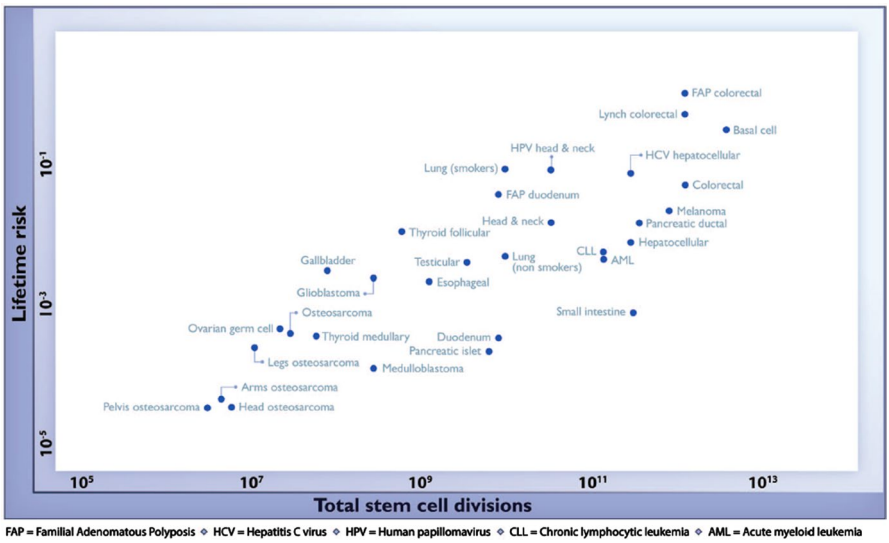

Fig. 1 The relationship between the number of stem divisions in the lifetime of a given tissue and the lifetime risk of cancer in that issue (From Tomasetti and Vogelstein 2015).

There are two parts to their argument. First, they argue that lifetime turnover of stem cells is associated regular acquisition of mutations, and that differences in number and rate of turnover of stem cells in different tissues and organs is a likely explanation for the significant magnitudes of difference in average incidence of different cancer types. Second, they claim that these "stochastic" influences can be quantified, and shown to be more important in accounting for differential variation in lifetime cancer incidence than either environmental or genetic factors.

Let us take each step of the argument in turn. They provide a striking image, representing "total number of stem cell divisions during the average lifetime of a human on the x axis and the lifetime risk for cancer of that tissue type on the y axis" The graph appears to show an almost linear relationship—the correlation is almost 1:1, extending across five orders of magnitude (Fig. 1).[5]

At first glance, this may seem to be a very poor argument. Surely Tomasetti and Vogelstein are not simply pointing to a correlation and asserting that it is a causal relationship? Are there independent reasons to think that stem cell turnover is the main cause of the difference in average incidence across cancer types?

There are independent reasons to think this. Tomasetti and Vogelstein are not simply noting a correlation and arguing that it is causal. Rather, they are drawing upon decades of research on basic cancer biology (concerning how mutations change the behavior of cells, endogenous mutation rate, patterns in average lifetime risk, stem cell research, etc.), which point to stem cell turnover as a major cause of

[5] Strictly speaking, "Spearman's rho = 0.81; $P < 3.5$ Å $\sim 10-8$) (Fig. 1). Pearson's linear correlation 0.804 [0.63 to 0.90; 95% confidence interval (CI)] was equivalently significant ($P < 5.15$ Å $\sim 10-8$)."

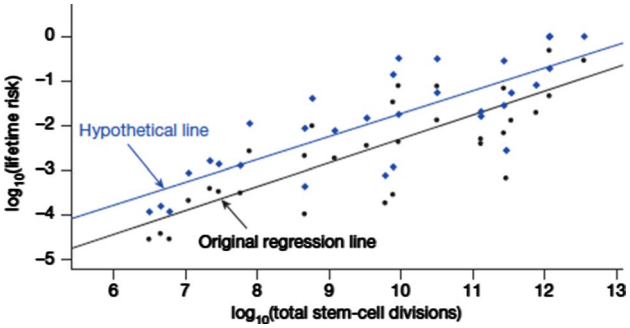

Fig. 2 Correlation analysis of stem-cell division and cancer risk does not distinguish contribution of extrinsic versus intrinsic factors to cancer risk. The black dots are data from Fig. 1 of Tomasetti and Vogelstein, and the black line shows their original regression line. The blue diamonds represent the hypothesized quadrupled cancer risks due to hypothetical exposure to an extrinsic factor such as radiation. The blue regression line for the hypothetical risk data maintains the same correlation as the original black line, albeit reflecting a much higher contribution of extrinsic factors to cancer risk (From Wu et. al. 2016)

these striking disparities in incidence. This part of their argument has the form of an inference to the best explanation, or, perhaps better, inference to a hypothesis (though it's far from clear whether this paper's argument is in the spirit of abduction "discovery" or "confirmation").[6] They claim that their favored hypothesis predicts and explains this pattern better than any alternative (or at least, any they consider). Indeed, no other hypothesis would seem to explain so well this striking variation in average incidence.

That is, the correlation is only one (of several) sources of evidence favoring their hypothesis. First, it's widely known that cancer incidence by and large increases as we age, and that cancer cells typically bear mutations to genes that make a causal contribution to the "hallmark" behaviors of cancer cells. These genes affect important regulatory pathways in the cell, detecting errors in replication, halting and initiating the cell cycle (Hanahan and Weinberg 2011). Second, it's widely agreed that some proportion of mutations are acquired at a relatively regular rate via somatic cell division over the course of a lifetime, and that somatic cell division rates vary across tissues and organs (Martincorena and Campbell 2015). And, (though this is somewhat more controversial) it's widely believed that different tissues and organs contain different proportions of somatic stem cells,[7] constantly renewing cells,

[6] There is a great deal more to say about IBE, whether it really is two forms of argument or one, whether all IBE begs the question, or instead can be represented warranted inference, such as by appeal to relative likelihood or Bayesian formal models of confirmation. I set these (endless) debate to one side here, but see Plutynski (2011) for a review of the history of these debates over IBE's form and warrant. See the "Appendix", however, for two different ways of formally reconstructing the argument. The latter draws on Schupbach's formal representations of IBE as a relative likelihood argument (2016). Many thanks to Jonah Schupbach for his feedback and suggestions.

[7] There is a rich philosophical literature on the variety of senses of "stem cell" in the biological literature, methodological challenges facing experimental work on stem cells, and on the role of cancer stem cells in cancer biology, which deserve attention, but which I cannot review in any depth here. However, I refer the reader to the excellent work of Fagan (2013), and Laplane (2015).

which give rise to cells that make up differentiated tissues and organs in the body. Effectively, all of the above are background assumptions, which, if true, render their hypothesis a likely explanation for the disparity of incidence across cancers.[8]

However, there are some researchers that dispute what is often called the "somatic mutation theory," the view that somatic mutations are causally responsible for the behavior of cancer cells and initiation of a tumor. The "tissue organization field theory" (TOFT) is a competitor to the somatic mutation theory (SMT) (Sonnenschein and Soto 2008; Soto and Sonnenschein 2011). On the former view, "carcinogenesis represents a problem of tissue organization comparable to organogenesis, and (ii) proliferation is the default state of all cells." (Soto and Sonnenschein 2005, p. 103) Soto and Sonnenschein have argued that Tomasetti and Vogelstein do not actively consider whether an alternative theoretical framework might equally well explain the patterns we see. There is some disagreement among both scientists and philosophers concerning whether TOFT is a replacement for (or perhaps supplement to) somatic mutation theory.[9] Nonetheless, many critics agree that Tomasetti and Vogelstein paid insufficient attention to further (and complicating) endogenous factors in cancer's etiology, and that this undermines the second stage of the argument (see section "Replies and controversies").

The second part of the argument asserts a different claim: namely, one can *quantify* the proportionate contribution of "chance" or "stochastic" factors to relative cancer incidence, and moreover, that this contribution is higher than any "deterministic" causes (which is the label they assign to environmental and inherited factors). This is where their argument goes off the rails, because to make such an argument, they have to quantify the contribution of endogenous mutations due to different number and rate of turnover of such cells of each tissue and organ. They use this value to estimate what they call the "extra risk score," (ERS), a measure of the extent to which "deterministic factors such as environmental mutagens or hereditary predispositions strongly affect their risk." That is, ERS is a measure of the relative contribution of "exogenous" factors to cancer incidence, which they take to be any "extra" risk above that expected, due to endogenous "stochastic" mutation events. This second stage of the argument is far more problematic. Quantifying relative contribution of "endogenous" versus "exogenous" cancer risk overall runs into a number of difficulties (as we shall see, in section "Replies and controversies"), but I will briefly describe how they attempt to do so.

Tomasetti and Vogelstein assigned an Extra Risk Score (ERS) for each cancer type, "the product of the $\log_{10}$ value of lifetime cancer risk, r, and its lifetime number of stem cell divisions, lscd:

[8] In the first of the two formalizations of the argument in the "Appendix", this background knowledge are premises 1–7.

[9] It is this author's view that there is a middle ground between these competing "paradigms." (Plutynski 2018, 2020) That is, one can take both tissue organization and mutation to play important roles in cancer causation. An integrative approach that attempts to draw upon multiple perspectives is preferred. See also Malaterre (2007), Marcum (2009) and Bertolaso (2016).

$$ERS = log_{10}\, r \cdot log_{10}\, lscd$$" (Tomasetti and Vogelstein 2015, Suppl., p. 10)

They conclude that "that the greater the absolute value of this product is, the smaller the evidence for the presence of any environmental or inherited factor acting on that tissue." (Ibid.). In other words, they effectively assigned a specific value to the relative contribution of "endogenous" and "exogenous" factors in cancer risk (as if these were independent). Moreover, they take any value over and above the "endogenous" base rate (that part exclusively due to mutations acquired during stem cell turnover) to be "exogenous."

They also used machine learning analysis (K-means cluster analysis) to distinguish "R" versus "D" tumors, (where "R" stands for "replicative," and "D" for "deterministic") based on their "extra risk score" (ERS): the extent to which "deterministic factors such as environmental mutagens or hereditary predispositions strongly affect their risk." They refer to tumors with relatively low ERS as replicative "because stochastic factors, presumably related to errors during DNA replication, most strongly appear to affect their risk." (2015, p. 80) Despite the fact that they acknowledge (deep in an "Appendix") that there is no established way to decide beforehand the "correct" number of clusters generated using this particular clustering algorithm, and that cancer types are really on a continuum, they use ERS to classify cancers into exactly two types. They then suggest policy guidelines for investing resources in primary versus secondary prevention, based on which type a given cancer falls within. If the ERS for a cancer is high (or in the "deterministic" cluster)—that is, if there is a high cancer risk of that tissue type *relative to its number of stem cell divisions*—then environmental or inherited factors play a relatively more important role in that cancer's risk (in their view), and thus, one ought to invest relatively more resources in primary prevention for those cancers. (Primary prevention consists of measures such as vaccines against infectious agents, or smoking cessation programs.) If the ERS for a cancer is low, then primary prevention is less likely to make a significant difference in overall incidence, and so one ought to invest more in secondary prevention (e.g., screening for early detection). Though they grant that screening could further reduce high ERS cancers as well, it will be most effective (they claim) for "R" (replicative) cancers. It is these stages of the argument that (perhaps not surprisingly) most raised the hackles of critics, though critics raised objections with each step of the argument.

Replies and controversies

There are many different kinds of response to the Tomasetti and Vogelstein paper; some question the quantity or quality of the evidence, others the logic of the inference itself. Strikingly, the sort of objection raised at least in part overlapped with disciplinary specialty, suggesting different pragmatic interests, methodological commitments, and thus also arguably standards of evidence for claims about causation and explanation, shaped these critiques. I'll begin briefly with some illustration of how disciplinary specialty and pragmatic interest shaped various objections, and

then turn to a brief overview of the most substantive objections, explaining which ones hit the mark.

By way of example of specialist interests, Gold (2017), takes issue with their claim about the "stochasticity" of mutations. Gold is writing as a specialist in the mechanisms (and outcomes) of mutagenesis. Very briefly, by way of background, it's widely agreed that some proportion of mutations acquired during somatic cell division have no or little effect. In cancer cells, such mutations that seem to play no role in the typical behavior of cancer cells are often called "passenger" mutations. In contrast, "driver" mutations are those to genes that increase the ratio of cell birth to cell death, or change the behavior of cancer cells in ways that contribute to the onset and progression of disease (Bozic et al. 2010). Gold (2017) takes issue with Tomasetti and Vogelstein's claim that the driver mutations are (as he puts it) "purely "stochastic." While he grants that there is a background mutation rate of random DNA polymerase misincorporation errors that occur during each cell division (estimated to be $7.6\times10^{-10}(\pm3.8\times10^{-11})$ per base per cell division (Tomasetti et al. 2013), Gold argues that a high proportion of mutations to specific tumor suppressor genes (APC and TP53) are not (as he puts it) "purely stochastic," but occur at a specific frequency that can be predicted. In particular, APC is typically deactivated by "nonsense (65%), frameshift (27%), splice site (5%) and missense mutations (3%)… If mutations were random, the nonsense mutation frequency at any of these codons should reflect the relative population of codons present in the APC gene." (Gold 2017, p. 38) However, they do not. He argues similarly for TP53, that some codons are mutated more frequently, and in particular, CpG sequence, together suggesting that both mutations are more likely due to common mutagenic pathways: "This mutation pattern is consistent with the kinetically slow, but not stochastic, hydrolytic deamination of 5-methylcytosine residues at specific methylated CpG sites." (Gold 2017, p. 37) In other words, Gold claims that we can predict with some probability which mutations of which types yield driver mutations, and indeed, can associate particular mutation types with mutagenic causes (Gold 2017, p. 38). In a follow-up paper, Tomasetti and Vogelstein comment:

> R mutations appear unavoidable now, but it is conceivable that they will become avoidable in the future. There are *at least four sources of R mutations in normal cells: quantum effects on base pairing, mistakes made by polymerases, hydrolytic deamination of bases, and damage by endogenously produced reactive oxygen species or other metabolites*. The last of these could theoretically be reduced by the administration of antioxidant drugs. (2017, p. 1333, italics added)

In the above, Tomasetti and Vogelstein seem to place far greater emphasis on the "avoidability" of this or that mutation or mutation type than on their relative *predictability*. That is, their concern is not whether one or another mutation type is more common, or could be predicted as more frequently contributing to cancer's etiology. By referring to mutations occurring during stem cell division as "stochastic," they did not intend to endorse the (rather strong) sense of "stochastic" criticized by Gold. Rather, the sense in which Tomasetti and Vogelstein intended such processes to be a "matter of chance" seems to be either that they are "unavoidable," or "intrinsic"

by products of processes of cell differentiation, division, and development. While Gold's claim that the relative frequency of some mutations over others, and the association of such mutations with mutagenic factors weighs against this, it's unclear from the above which mutations and how many of them they take to be relatively unpredictable. Moreover, while quantum effects on base pairing might well be modeled as a perfectly random sampling process, other causes of mutation they list yield certain classes of mutation with higher or lower probabilities. It's notable that this comment fails to address Gold's concern: namely, that mutagenic factors are likely indicated in the CpG mutations to several tumor suppressor genes that play major roles in several cancers.

Another example of disciplinary specialty yielding different kinds of critique is as follows. Some critics disputed the value of Tomasetti and Vogelstein's question altogether. In what might seem initially to be a striking non sequitur, several epidemiologists point out that:

> … from a public health perspective the key issue relates not to which cancers are more common or rare, but rather to what is the overall burden of cancer, how much of this is preventable in principle, and how much with current knowledge? The difference between the two is an indicator of the need for further research to identify causes of cancer modifiable at the population level. (Davy-Smith et al. 2016, pp. 605–613)

That is, these critics seem to be saying that the matter of why some cancers are so rare and why others are so common is simply uninteresting. Of much greater interest is *which cancers we might reasonably prevent*. From the perspective of public health scientists, it may indeed seem to matter very little why some cancers are so comparatively rare and others so common, unless investigating this might yield better understanding of how to reduce the overall burden of cancer. Perhaps (unfortunately) this is exactly why Tomasetti and Vogelstein go beyond what is warranted in their initial paper, assigning cancer types to different categories, based on their "extra risk score."

By way of a third example of disciplinary specialty shaping critique, two statisticians (Sornette and Favre 2015) argue that Tomasetti and Vogelstein make a simple error in statistical reasoning. Namely, they fail to rule out the possibility that within any particular cancer type (e.g., small cell lung cancer) subpopulations of patients could have differential vulnerability, for instance, due to environmental and genetic causes unique to such subgroups. Such a variation in risk within subgroups is compatible with the empirical evidence of a strong correlation between the total number of cell divisions and average incidence of each cancer type, but might be erased by the kind of analysis with which Tomasetti and Vogelstein are engaged. Exogenous factors or genetic traits could impact both the cancer risks related to stem cell divisions and those that seem unrelated to stem cell divisions. Tomasetti and Vogelstein make a simple error in statistical reasoning by presupposing that these correlations are unconditional, and do not consider how conditional correlations might lead to the same pattern.

In a similar vein, Weinberg and Zaykin (2015), also point out that in their attempt to quantify relative contribution to overall incidence, Tomasetti and Vogelstein make

several errors in reasoning. First, by focusing on aggregated data and differences in relative incidence across cancer types, they obscure potential causes of overall incidence. Weinberg et al., make an analogy with traffic fatalities: were policy-makers to focus only on state-by-state incidence of such fatalities, they may be mistakenly led to believe that the major cause of such fatalities is miles driven per capita, rather than drunk driving. For, when comparing states, (70%) of the variation in log risk (per year) of a fatal crash appears to be explained by the log of the average number of miles driven per year per person, whereas drunk driving appears to only account for 0.02% of the fatality rate across states). But, when considering overall incidence, "fatal accidents disproportionately involve drunk drivers (with fractions varying from 16% in Utah to 44% in Montana)." (Weinberg and Zaykin 2015, p. 3)

Moreover, they point out how these causes are simply not independent in the case of cancer, complicating any attempt to partition causal contributors to overall risk:

> … environmental exposures, germ-line genetic variants and random events like replicative errors typically act in concert; the effects cannot be treated as separable. It is a mistake to assume that one can partition etiologic factors into contributions that sum to 1.0, as in the notion that two-thirds of cancers are due to bad luck and therefore at most one-third could be due to environmental and inherited genetic factors. Because of joint effects, contributing causes often have attributable fractions that add to more than 1.0. The intellectual disability syndrome secondary to phenylketonuria is a well-known example where the fraction attributable to genetics is 1.0, while the fraction attributable to environment is also 1.0, because the outcome requires both a dysfunctional metabolic gene and an environmental exposure (dietary phenylalanine). (Weinberg and Zaykin 2015, p. 3)

In other words, Weinberg et al. seem to be suggesting that Tomasetti and Vogelstein are assuming that one can partition the relative discrete contribution of endogenous and exogenous factors to cancer overall, by relying on "attributable fraction" assuming that it must sum to one. But, this ignores the fact that the causes in question are not independent—i.e., cancer incidence is a joint effect of both exogenous and endogenous factors.

There were many more criticisms of this paper in the literature, but for the sake of economy, I'll summarize the three more substantive ones that were mentioned more than once in the literature, before moving on to a consideration of some larger methodological upshots in the next section:

- First, several critics pointed out that the options they consider are not mutually exclusive.
- Second, they do not take into consideration additional endogenous causal factors that (arguably) make a systematic difference to relative cancer incidence, and may well remove the puzzle we began with.
- Third, they exclude some cancers for which we have little data on stem cell divisions; and, perhaps more seriously, there is reason to question the data they rely upon for cancer stem cell division in the first place.

Let us consider these objections in turn. First, Wu et al. (2016) (epidemiologists) make an elegant version of the first objection by appeal to a thought experiment, which is accompanied by a vivid image (Fig. 2):

> … consider a hypothetical scenario of a sudden global emergence of a very potent mutagen, such as a strong radiation burst from a nuclear fallout, which quadruple the lifetime risks for all cancers. In this scenario, it transpires that the proportion of cancer risk caused by intrinsic random errors would be small (at most on quarter if we assume all of the original risk was due to intrinsic processes) However, if we conduct regression analyses on either the new hypothetical cancer risks or the current cancer risks as reported, against the number of stem-cell divisions, the correlations from both cases would be 0.81. This thought experiment negates the ability of the correlation to detect solely the contribution of intrinsic factors as it cannot distinguish between intrinsic and extrinsic factors. (Wu et al. 2016, p. 44)

As this vivid image makes clear (Fig. 2), it's certainly possible that overall rates of cancer have risen—such that while on average there is a correlation of cancer rates in different tissue types with rates of stem cell turnover, the most significant contributor to cancer overall is environmental. This is an argument against Tomasetti and Vogelstein's claim that ERS scores enable them to estimate contribution of stochastic versus exogenous causes to "overall" cancer incidence. Wu et al. (2016) use their thought experiment to argue that exogenous factors could well be the more significant causal contributor to overall cancer risk, particularly in the developed world.

This is a sound objection, and Tomasetti et al. make an effort to address it in a subsequent publication (2017), where, drawing upon cancer incidence data from 69 countries (including both developed and developing countries), they demonstrate that: "the correlation between cancer incidence and the number of stem cell divisions in various tissues cannot be explained by peculiarities of the U.S. population or its environment. This correlation is observed worldwide, as would be expected for a fundamental biological process such as stem cell divisions." (Tomasetti et al. 2017, p. 1333)

Whether one finds this argument persuasive or not, there is a second concern that is perhaps equally if not more serious. Several authors point to a variety of endogenous factors apart from stem cell turnover that potentially play important roles in shaping relative rates of incidence in cancers in different tissues. Such factors could well explain (or explain away) the puzzle that Tomasetti and Vogelstein begin with. That is, the differences in incidence could be explained not only by different numbers and rates of turnover of stem cells, but also a combination of variation in organ size, variation in tissue organization and patterns of clonal expansion, variation in the typical number (and type) of target mutations associated with cancer in different tissues, different levels of apoptosis, or different levels of immune surveillance across tissues and organs (Nunney and Muir 2015; Noble et al. 2016; Nowak and Waclaw 2017). In sum, Tomasetti and Vogelstein focused on only one of many potential endogenous causes of cancer— but, all of these, in some combination, contribute to relative cancer incidence. Moreover, all these endogenous factors are

products of evolutionary history; i.e., there was likely selection for or against various tissue or organ-specific mechanisms that protect against cancer. This in turn likely contributes to a degree of redundancy and plasticity at the cell and molecular level, such that accumulation of mutations can occur without necessarily yielding a cancer. That is, not all mutations in cancer cells play the same role in cancer initiation and progression; some "constrain" or "control" the activities of others (Bechtel 2018).

As several evolutionary biologists point out, endogenous variations in a variety of factors (e.g. from variation among mutations and their effects to organ size), across tissues in and organs likely led to selection for more or less modifying factors in risk across tissues and organs (Nunney and Muir 2015). Nowak and Waclaw put the point succinctly:

> … evolution could have generated additional checkpoints—for example, requiring a larger number of subsequent driver mutations (those that spur cancer progression), or better immune surveillance, in tissues that have more stem cell divisions… there may be, for reasons of tissue geometry…, a smaller number of effective stem cell divisions, and only such cell divisions contribute to the risk for cancer initiation" (2017, p. 1267)

To sum up, stem cell division is one, and only one of several constraints on the initiation and progression of cancer. Once one considers the long evolutionary history we have been subject to, one cannot simply take any deviation from expectation based on the number of stem cell divisions to be driven by "exogenous" causes. Though lifetime patterns of stem cell division surely affect lifetime risk, the consequences of these differences have themselves shaped the evolution of checkpoints at the genetic and tissue organization level that both prevent and promote cancer. Such factors act in concert, controlling the effects of mutation on cancer initiation (Nelson and Bissell 2006).

Moreover, the model of stem cell turnover and somatic mutation Tomasetti and Vogelstein seem to be relying upon more or less takes for granted that mutation events are probabilistically independent; but some mutation events acquired early on can have "catastrophic" downstream effects, accelerating the onset of cancer by permitting or enabling yet further mutation events that increase rate of progression to disease. The presence and prevalence of such chromosomal changes of major effect varies across tissues and organs (Gröbner et al. 2018). Arguably, taking into account these this further "endogenous" risk (since the role of "catastrophic" mutation events varies across tissues and organs) would significantly complicate the estimate of relative contributions to endogenous "intrinsic" cancer risk.

In sum, relative cancer incidence in any tissue is a product of a complex developmental and evolutionary history, yielding distinctive patterns of not only rates of somatic cell division, but also levels of apoptosis, immune surveillance, and much else. So, one cannot simply pin any variation from *expectation based on stem cell turnover* on "exogenous" causes; the extent of mutation, as well as the effects of mutation, are mediated by a variety of additional endogenous and exogenous factors in interaction. Such interactions between these factors complicate attribution of an effect to one or another cause. Both Tomasetti and Vogelstein and some of their

otherwise apt critics (Wu et al. 2016) seem to fall prey to this oversimplifying move, as Nowak and Waclaw point out.

There are a wide variety of ways that the picture of mutation accumulation that Tomasetti and Vogelstein have relied upon have been empirically challenged. First there appears to be ample evidence of hypermutation in cancer (Roberts and Gordenin 2014)—that is, mutation events that led to effectively a massive or radical transformation of the genome at some stage in cancer progression. Second, there is a great deal of mutational heterogeneity in human cancers, or variation in the number and type of mutations typical of cancers of the bone, brain, skin, etc., suggesting that the processes of mutation acquisition shaping cancer in the brain, for instance, may well not be comparable to cancer of the skin, etc. (Salk et al. 2010). Third, aging and acquisition of mutational variation across tissue is normal, and need not yield cancer, suggesting that their particular model of mutation acquisition as a simple, stepwise process (unmediated by, e.g., tissue organization) is flawed (Martincorena and Campbell 2015). And, fourth, some cancers are associated with distinctive chromosomal aberrations, such that only one or two specific changes can yield a cancer (Duesberg et al. 2000).

The third and final objection has to do with the quality of the evidence they rely upon the correlation itself. Several epidemiologists argue that Tomasetti and Vogelstein omitting several cancers from their analysis complicated their analysis, as did drawing their data entirely from the U.S. population. And, stem cell researchers contested their estimates of the number and rate of stem cells (Clevers 2018).[10] This is indeed a serious objection; as yet, we know very little about number and rates of stem cell division. This is an active area of research; in other words, the very data the authors are relying upon to yield the correlation they identify are contested.

What is striking in this discussion is that each discipline came at this paper from a different perspective: focusing either on failures regarding inclusion or exclusion of relevant causes, or quality of evidence. A larger issue in this debate is the matter of when we are warranted in reasoning from effects to causes, or more precisely: whether (and when) one is warranted in inferring from population level effects to relative causal contributions of different (and complex, interacting) population level causal variables. The inevitable empirical underdetermination involved in such inferences is confounded by the problem that events at a variety of temporal and spatial scales are acting an interacting in ways that make separating out relative causal contribution to lifelong relative cancer risk in different tissues and organs enormously difficult. Are there any methodological upshots we can draw from this case?

[10] Indeed, as Batlle and Clevers (2017) has argued, our biological understanding of "stemness" itself is a moving target. Whether and how cell hierarchies present in tumors are organized, such that they yield higher or lower rates of cancer, is still not well-understood, given that our experimental work on this question is indirect. Variation in tumor niches may differentially affect whether and how often tumor cells acquire the stemness phenotype (see also: Laplane 2015; Laplane and Solary 2019).

Diagnosing the controversy: causation, chance, explanation and control

So what exactly went wrong in the "bad luck" debacle? There are (at least) three key issues worth highlighting in this case:

- First, the authors used the term "chance" and its cognates in a variety of different senses, sometimes to refer to subjective states of affairs and sometimes to what philosophers call "objective" chances.[11] This arguably confused their critics, leading to objections that both missed the mark in some cases, and were warranted in others.
- Second, the critics brought different interests, and relatedly, different methodological criteria (or, standards of evidence) to bear on the debate, which informed their criticisms of both the framing of the options to hand, and the relevant causal variables.
- Third, the authors made some inferences that were weak at best, and deeply flawed at worst, such as treating two options as mutually exclusive and conditionally independent that were not, failing to consider all relevant evidence, such that they ignored alternative possible (and indeed, arguably, plausible) explanations, and moving from a claim about causation to a judgment about potential intervention that went far beyond what the evidence warranted.

In sum, there were legitimate concerns with not only how talk of "chance" was confused and confusing, but also with the logic of the argument, and the empirical evidence considered. Given the complex interactions between endogenous and exogenous factors in cancer, assigning a relative causal role for each (let alone assigning endogenous factors the label "chance") was criticized by cell and molecular biologists, public health scientists, and statisticians. That said, it's not impossible (in very specific circumstances) to make warranted arguments for the relative strength or weakness of this or that causal contribution to a population level pattern (whether in disease incidence, or in relative survival), even when such variables are not unconditional contributors to such effects. At very least, one should grant that there are more or less well-supported such inferences (Hill 1965; Haldane 1964). Otherwise, it would be rather difficult to investigate environmental causes of ill health, or evolutionary causes of changes in populations. Below I briefly consider the relevant factors that ought to govern assessments of evidence for such claims of causation with respect to population level outcomes.

[11] Regarding whether objective chances (or all talk of probability) can be reduced to propensities or relative frequencies, I do not think it's necessary for me to take a stance in service of diagnosing the conceptual and methodological confusions at issue in this debate. However, I endorse a broadly pluralist view about probability, akin to that defended by Suárez (2020, 2017), and am a realist about the objectivity of macro-level probabilities. If they can be measured, are robust, and predictive, they are objective, whether they are multiply realized or reducible (see, e.g., Sober 2010 for a discussion of the "reality of macro-probabilities" in evolution, for an analogous case).

The first step in any inference to a hypothesis about population level outcomes (such as cancer incidence), is to make the contrast between competing hypotheses explicit, and to assure that they are mutually exclusive. For instance, Doll and Hill's (1950, 1952, 1954) argued that cancer incidence (and lung cancer incidence in particular) was higher among smokers than non-smokers. In this case, the relevant alternative causes are exclusive options: smoking and not smoking. They found that there was a strong dose response relationship; the more an individual smoked, the higher their risk. In addition, they had independent evidence to suggest that smoking was a significant contributor to increase in incidence of cancer; tobacco tar had been identified as a carcinogen in animal studies, smoking was known to irritate the lungs, and the lungs of lifelong smokers were characteristically damaged. They also considered a variety of other potential causal variables (pollution in urban versus rural areas), and were unable to identify any of these as having as consistent and strong an effect as smoking. Hill (1965) describes several factors of relevance to assessing such inferences regarding the role of environmental exposure in disease: strength, consistency, specificity, temporality, biological gradient, plausibility, and coherence. In the case of Doll and Hill's original paper: the association was quite strong (smokers are between 10–20× as likely to be found among those with lung cancer).

Subsequent studies to the original (1950) paper (1952, 1954) showed that the association was consistent across a variety of background conditions and in different populations of patients. Of course, smoking occurred prior to the outcome, and there was a consistent direction of increase in incidence among longer term and more frequent smokers. Moreover, there was a plausible biological explanation for how smoking contributed to cancer risk, coherent with what was then known about the mechanisms associated with cancer. Though a good deal had yet to be learned about the detailed mechanistic basis of cancer, to the extent that there were well understood mechanisms, smoking seemed to be a plausible exposure likely to lead at very least to inflammation, if not also to mutagenic toxins, both known causes of cancer over the long term. The inference was coherent with other scientific research and conclusions. As mentioned above, toxicological studies had demonstrated that exposure to tobacco tar in animals over time could lead to the development of tumors. In the case of such inferences, the evidence strongly supported their hypothesis.[12]

In contrast, Tomasetti and Vogelstein may have had weak though plausible grounds to infer to their hypothesis, at least at the first stage of the argument. However, they attempted to offer not only a hypothesis about stem cell turnover as a major cause, but to assign it a specific quantitative measure. Worse still, by tying their argument regarding etiology to the claim that ERS (extra-risk scores) could provide guidance in allocating resources to preventive measures of various sorts,

[12] One could well argue that Doll and Hill's evidence was more than sufficient to establish that smoking caused lung cancer by 1950, but "sufficiency" in my view depends upon what one wishes to use such causal information for. We can have better and worse evidence, and better or worse reasons to regulate (e.g., saving people's lives versus making a profit). There is of course a great deal more to say here about the pragmatic and value-laden character of inference in epidemiology and public health (see, e.g., Reiss 2015; Broadbent 2011a, b; Plutynski 2018).

they went far beyond what the evidence warranted. The second stage of their argument was flawed even in principle; even if they could pull apart endogenous and exogenous etiological contributors to relative cancer incidence, ERS (alone) would not predict relative effectiveness of preventive measures.[13]

In contrast, consider Haldane's defense of "beanbag" genetics, against Mayr's (1963) attack, namely that "To consider genes as independent units is meaningless from the physiological as well as the evolutionary viewpoint." Essentially, Mayr is suggesting that the exercise of developing mathematical models that treat the processes of selection as if acting on genes unmediated by development, and without attending to organisms' complex interactions with their environment, is "meaningless." Haldane points out that beanbag geneticists can grant that genotype and phenotype are mediated by complex causal pathways, and that evolution is not mere selection of "beans" from a bag (selection does not act on traits in isolation), but involves selection acting on the organism as a whole. Nonetheless, Haldane argues, mathematical population genetics gives one a way to describe, explain, and test out alternative possibilities for the major causes of change in populations: "One must try many possibilities before one reaches even partial truth…" However, he continues with the following warning: "There is, however, a danger that when a mathematical investigation shows a possible cause of a phenomenon, it is assumed to be the only possible cause." (Haldane 1964, p. 351)

This lesson may well have been useful for Tomasetti and Vogelstein to keep in mind. As we've seen, just because a variable predicts the data, it does not explain the data. What more then is required? At minimum, the candidate cause needs to temporally precede the effect. We should also have independent reasons (apart from correlation) to think the cause makes a difference to the rate at which effect occurs. Tomasetti and Vogelstein did have such reasons; cancer, as we've seen, is a product in part at least of mutations, and mutations accumulate at a regular rate with stem cell division. So, there were in principle good reasons to consider this a candidate hypothesis for the outcome they were seeking to explain. That said, they did not consider how rate and number of stem cell turnover shaped (and was shaped by) other endogenous factors, as well as how causal interactions between endogenous and exogenous contributors could systematically affect relative cancer incidence. That is, stem cell turnover was certainly not the exclusive "endogenous" cause of cancer, nor was it "discriminate" in its effects.

Put more generally, whenever two hypotheses are taken to be competing alternative hypotheses, one needs to be able to say that (and how) they each make a difference to the outcome, in a way that can discriminate their relative contribution. In this case, it was not clear that "endogenous" factors such as those described were legitimately contrasted with "exogenous" factors, because the two were simply not independent of one another. They combine in ways that make systematic differences

[13] As they later acknowledge (2017), etiology is an independent matter from relative effectiveness of primary and/or secondary prevention. That is, whether or not a primary or secondary preventive policy is warranted has far more to do with whether the measures in question are practically effective than whether they ought to be so in principle.

to average incidence across tissues and organs, and likely in different ways across different tissues, given the long evolutionary history to which humans (or species with sufficiently similar functional organization of tissues and organs) were subject.

That a variable makes a "systematic" difference is a vague idea, but can be made somewhat more precise. Woodward's (2010, 2018) suggests that there are (at least) two criteria in assessments of the adequacy of causal explanations in biology, which seem to be of relevance here: stability and proportionality. Stability of causal relationship "has to do with whether it would continue to hold under changes in background conditions," and proportionality, "with the extent to which a causal claim fully captures conditions under which variations in some phenomenon of interest occur" (Woodward 2018). Put intuitively, in considering which causal variables to compare and contrast, we want to know whether the candidates are "stable" in the sense that we expect them to make a difference to the outcome in question, even if many background conditions were different. In other words, we want a candidate cause to be relatively invariant in its effects. Moreover, we wish to give explanations that appeal to causal variables that are described in such a way that "there is a match between the variation in the cause and the variation in the candidate effect." Or, proportional variables should "capture the full range of dependency relations that hold in the situation of interest without falsely implying dependency relations that do not hold" (Woodward forthcoming, p. 11).[14] Key here is the idea of a "situation of interest." Woodward does not take failures of proportionality to be instances of false causal claims. Rather, they are criteria for variable choice that are optimal, and are matters of degree, not kind:

> … a choice of variable V_i (and of the dependency claims regarding E in which V_i figures) satisfies proportionality better than an alternative choice from **V** to the extent that those dependency claims satisfy *Falsity* and *Omission* above—that is, to the extent that (i) non-existent dependency relations involving E are not falsely represented as present (as noted earlier this can be understood in terms of satisfaction of **M**) and to the extent that (ii) existing dependency relations (from among the variables in **V**) involving E are represented. When we have specified a cause variable and associated dependency relations delineating the conditions under which all possible values of E occur, we have fully satisfied **P**. (Woodward forthcoming, p. 277)

That is, identification of stem cell turnover as a major cause of differences in relative incidence was not—per se—a false claim. What was false is the treatment of this as the sole or most important causal variable, and certainly with respect to outcomes of interest to public health scientists. Failures in stability and proportionality seem to be at play in several ways in critics' responses to Tomasetti and Vogelstein's

[14] In this sense, statistical explanations in biology can be said to be "autonomous" from those at other temporal or spatial scales. As might be expected, then, there is a debate about Woodward's virtues of stability and proportionality, and whether they indeed favor "higher level" explanations. While considerations of space prohibit addressing these matters at length here, see Shapiro and Sober (2012); Weslake (2013); Franklin-Hall (2016), and Woodward's reply (2018, forthcoming).

reasoning and argument. That is, critics argued that Tomasetti and Vogelstein failed to reckon with the possibility that subclasses of cases within any particular category of cancer were largely due to factors apart from stem cell turnover. In this way, existing dependency relations were not represented. Indeed, biologists complained that, many endogenous factors that contribute to cancer risk apart from stem cell turnover were simply ignored. This too was a failure of omission or lack of relevant information, not a strictly speaking false claim.

By assigning stem cell turnover a specific quantitative value, however, and suggesting that this was directly relevant to prevention, Tomasetti and Vogelstein were not simply making an error of omission, but a false claim; that is, they were also simply engaged in a misrepresentation. What methods of intervention will be effective is due to factors largely outside the control of biology: for instance, whether or not addictive behavior is associated with a given environmental carcinogen, which screening methods are more or less sensitive or specific, whether patients will adopt a screening method, or what kinds of follow up screening and treatments are available. More generally, if we care about modifying cancer incidence, what we need to understand is not only which causes are "more" significant, but which can be effectively manipulated, and how. No doubt, endogenous and exogenous causes are often so intertwined that attempting to pull apart their relative contribution and assign a relative value to each is problematic for a variety of methodological reasons, as several statisticians pointed out. More importantly to many public health scientists, it was beside the point; what counts is identifying which risk factors are manipulable, how, and how much.

On how cancer is a matter of chance

One may be left wondering, however, what—if anything—this debate teach us about the role of "chance" in cancer.[15] In what senses—if any—can one meaningfully speak of cancer as a matter of "chance"? As mentioned above, "chance" is predicated of an event, population of events, or a process, in a variety of ways in biomedicine:

- equiprobable with the relevant alternatives,
- due to causal processes that are not determinate, or only yield distinct outcomes with some probability,

[15] Many philosophers argue that the causes that matter are the "ultimate" or "foundational" ones—those that are the concerns of physicists. If indeed the fundamental sciences are deterministic, then nature as a whole is deterministic. On this view, the question of whether cancer is a matter of "chance" or luck is simply ill-conceived. The problem with such arguments is that they trade on matters that no empirical evidence so far can decide. Whatever you make of a priori arguments for causal determinism, the jury is still out for those of a more empiricist bent. Even our best physical theories leave open the question of whether causal determinism is true (Hoefer 2016). Moreover, it's in principle possible that regularities at the macrolevel—or the domains that concern us with respect to explanation and prediction in biology—are only weakly constrained by micro-scale regularities (see, e.g., Ismael 2016, 2017; Batterman 2011).

- due to species-wide ways in which design fails (in this case, attributable to the distinctive functional organization of each tissue and organ and vulnerabilities associated with such distinctive organization),
- due to causes we as yet have no means to control, or are unlikely to develop means to control.

Which of these is at play in cancer? First, where, when, and how often, various genes are mutated over the course of a lifetime, during somatic cell division, is at best a matter of probabilities. As we've seen, the picture of how and why those probabilities yield the outcomes we see is more complicated than Tomasetti and Vogelstein suggest. They are conceiving of the process by which mutations are acquired as a "stepwise," one where each mutation is a relatively independent event of downstream mutations, and cancer comes about given a certain number of mutations are acquired. But different kinds of mutation have different downstream effects. If a mutation occurs early on in a gene that plays an important role in either the process of replication itself, or correction of errors in replication, then it can accelerate cancer onset. Some can yield significant disruptions to chromosomal duplication, in turn serving as promoters of yet further mutations. So, it's simply not true that relative cancer incidence is anything like a direct effect of rate of turnover of stem cells.

Of course, what Tomasetti and Vogelstein label "deterministic" factors in the promotion of cancer are also a matter of "chance" in the sense of objective probabilities. That is, families with specific inherited mutations (to genes, e.g., TP53, in Li Fraumeni syndrome, or BRCAI and II in breast cancer) have higher probability of developing cancer (and at younger ages). Not all those bearing such mutations will get cancer. Likewise, environmental exposures governing cancer incidence are probabilistic (or "chance") factors, in that exposure does not fully determine, but only increases the probability on average (often with a dose–response relationship). We know ways to intervene on such exogenous factors (if not in practice, at least in principle), and we can predict the direction of their outcomes with probabilities that are empirically well-established, at least within some range.

Tomasetti and Vogelstein's focus was on probability of a mutation per cell division, but as Gold argued (above, 2.1), at a finer grain, there are different kinds of mutation, with different probabilities of occurring. And, different kinds of outcome of such mutations occur with different objective probabilities as well. There are objective measures of the probability of a base pair mutation occurring per somatic cell division as well as objective measures of probabilities of higher rates of mutation in some regions of the genome, and in certain codons, rather than others. There are also higher rates of mutation in some populations versus others (males versus females) given different initial conditions (rate of growth affects rates of error in somatic cell division). And, there are differential effects of various epigenetic modifications to the genome over the short or longer term (e.g., there are differences in telomere shortening in different groups, that are probabilistically associated with higher rates of cancer). Mutation is thus a "probabilistic cause" in the sense that whether or not it's "ultimate" basis is indeterministic, we can assign probabilities to whether and how often one or another type of mutation is likely to occur, in this or that somatic cell, and what effects might ensue.

We also have good explanations for how and why mutations occur as often as they do, where they do, on the evolutionary time scale (Kimura 1968), and during the lifetime of the organism (Nunney 2018; Noble et al. 2016). These are objective probabilistic measures, and can be linked to cancer incidence in different species or subpopulations, given different evolutionary history, ecological conditions, and life-history trade-offs (Aktipis et. al. 2015). The somatic mutation rate we find in a given population of cells dividing over a lifetime is a product of these trade-offs between the selective advantages of accurate replication, and the cost of error in cell division over the course of a lifetime (Kimura 1968). All this occurs at the "macro-scale", and such processes are evolved to "screen off" any variation at the micro-level that makes significant differences to fitness over the lifetime. This is exactly why we don't get cancer more often, and indeed, can ignore micro-level changes for the most part, with respect to macro-level traits that affect fitness, such as cancer. Though, of course, there are likely many other contingencies or probabilistically relevant causal factors at this evolutionary scale; the random extinction of this or that lineage could have eliminated variants with slightly different developmental processes, yielding different organ or tissue size or processes of differentiation, which in turn could have contributed to higher or lower risk of mutation, and of cancer.

In sum, there are a variety of ways in which "chance" has been at work in relative cancer incidence, at a variety of temporal and spatial scales. Of course, by "chance" here, I mean objective probabilities; we can, for most purposes in science, set aside the question of whether these probabilities are "fundamentally" indeterministic. That said, sometimes scientists do assume that at base, there is some relatively indeterministic cause at work, (quantum effects, as discussed in section "Replies and controversies"). When scientists speak of an outcome or process as "due to chance," they are not always precise about whether they intend it in the "subjective" or "objective" sense. The best one can do is gather the intended meaning by attention to context. As we've seen, a key element of such contextual analysis is attention to the target of explanation—or, the phenomenon to be explained. The phenomena cancer scientists seek to explain are typically types of event or processes (e.g., mutation types, typical processes of invasion or metastasis in liver cancer), or population level measures of various event-types (e.g., lifelong average incidence of cancer in women in the U.S., lung cancer mortality in the U.S. between 1990–1999). To the extent that we can specify an outcome, and treat competing causes as relatively independent, stable, and proportionate contributors to that outcome, we are warranted in measuring their relative contributions, and this measure is typically given in terms of probabilities, or more precisely, ranges of probabilities.

Unfortunately, once we consider cancer incidence in general, such inferences become rather more complicated. Causes at a variety of temporal and spatial scales vary in ways that lead to variation in such "coarse grained" outcomes. Such processes are by and large conditional in their effects. Thus, one cannot read causes directly off of coarse grained probabilistic associations, because they supervene over many causal interactions that together yield patterns and processes of caner incidence, in ways that are often highly context-dependent. That is, all of the below are

acting and interacting with respect to overall cancer incidence, and we cannot treat them as strictly independent:

- Genetic mutations.
- Epigenetics (e.g., hypo and hypermethylation affect which genes are expressed, and thus how or whether cancer phenotype comes about)
- Organ or tissue size, or organism size (e.g., due to sex differences or species averages).
- Stem cell turnover (affects rate of acquisition of mutations).
- Rate of growth/type of growth over the course of development (e.g., sex specific variation in rate of growth affects, for instance, rate of childhood brain cancer).
- Immune function (e.g., sex specific sensitivity of immune function or dysfunction can affects sex-specific rates of cancer, inflammatory properties (e.g. bone versus lung tissue) can affect tissue or organ specific cancer incidence).
- Environmental exposure, such as infection (e.g., HPV), smoking, or UV radiation.
- Sex-specific effects (e.g., role of hormones in breast, ovary, and prostate cancers, & associated role of developmental stage, parity, life history traits in exposure).
- Age-specific effects (over time "checkpoints" on cancer break down in ways that predictably yield higher average risk as we age).

These causes act and interact together to yield patterns and processes of cancer incidence, progression, and mortality, in highly context-sensitive ways. A mutation event that might yield a rapid progression to disease in one subset of the cell population that makes up the skin might have no effect in the breast or lung. Whether various outcomes come about, is at best, something we can predict with some probability given changes in these variables. Such prediction and explanations are typically highly local. Even though we know a good deal about how these very complex causal processes yield cancer, there is a good deal of empirical underdetermination about the contribution of each such variable to any particular outcome. Generalizing to the relative contribution of, e.g., inherited v. acquired mutations to various cancer types and subtypes can be quite difficult (Hu et al. 2016).

In sum, assigning anything like a specific value to the relative contribution of "luck" or "chance" (in Tomasetti and Vogelstein's sense) is problematic, because stem cell turnover is neither the exclusive, nor unconditional, endogenous factor at work in relative cancer incidence. Put differently, it is not a stable and proportionate choice of variable, but highly contingent in its effects, in ways that vary across tissues and organs. Arguably, the prospect of resolving such debates with anything like precise measures of the relative role of "chance" is "gloomy" (paraphrasing Plomin and Daniels 1987, on the role of genes in behavior).

Conclusion

I have argued that first, Tomasetti and Vogelstein's talk of "chance" or "luck" was used to refer to classes of event that were not distinct in kind from so-called "deterministic" causes, in that both can be assigned objective probabilities. The only meaningful way in which the former class of causes differs from the latter in being "chancy" is in being due to species-wide ways in which design fails, or "endogenous" causes we as yet have no means to control. This inconsistency in talk of "chance" led to confusion in the literature. Second, the debate was complicated by the fact that different researchers had very different pragmatic interests, and thus different assessments of relevant causal variables "proportional" to outcomes of interest. Last but not least, however in principle plausible the causal link between stem cell turnover and relative incidence, the authors should have stopped with claims of "possibility." As Haldane argued, "a danger that when a mathematical investigation shows a possible cause of a phenomenon, it is assumed to be the only possible cause."

As we have seen, assigning anything like a specific value to the relative contribution of "luck" to relative incidence of different cancers is problematic on a variety of counts. There are contexts where it is possible to assign a value to relative causal contribution of exogenous variables to a population level outcomes. Hill and Doll's inference to the claim that smoking was a significant contributor to the increase in incidence in lung cancer is one such case. Unfortunately, Tomasetti and Vogelstein's attempt to quantify "chance" factors, and contrast them with "deterministic" factors, did not meet the minimal criteria for identifying and distinguishing competing causal variables.

Evidence for such judgments can be better or worse, and ideally should be founded on solid background knowledge. To some extent, the authors did rely on background knowledge that was well-established; the problem was that they reduced relevant endogenous causal variables to a single one. This allowed them to make a very simple predictive model. It is surely true that insofar as cells divide, mutations happen, and we are composed of cells that divide over the course of our lifetime. With age, the chance of cancer increases; and, different tissues and organs have different rates of turnover of cells. Not surprisingly then, different tissues and organs have different lifetime incidence. However, the relationship between the one and the other are not one-to-one. There are many moderating influences on the mutations acquired during stem cell turnover; endogenous causes of cancer interact in complex ways with one another and with exogenous factors. So, while we are all subject to trends and outcomes that to a large extent are matters of probability, the fact of the matter is that the causal factor they isolated and identified with "chance" was but one of many "chance" factors in cancer.

Appendix: formalizing Tomasetti and Vogelstein's IBE

First version

Premise 1. Genomic changes occur simply by chance during DNA replication.
Premise 2. The endogenous (somatic) per base pair mutation rate of all human cell types appears to be nearly identical.
Premise 3. There are a large number of somatic mutations known to exist in cancer cells.
Premise 4. Among such mutations are mutations to genes that make a causal contribution to the "hallmark" behaviors of cancer cells (e.g., oncogenes and tumor suppressor genes. The mechanisms by which these genes affect regulatory pathways in the cell, e.g., detection of errors in replication, or halting and initiating the cell cycle (cell birth and death) are in many cases well-understood (Hanahan and Weinberg 2011).)
Premise 5. Cancer incidence by and large increases as we age.
Conclusion 1. Somatic (i.e., acquired, rather than inherited) mutations are causally responsible for the initiation of a tumor. (from 1 to 5).
Premise 6. Stem cells—those cells in a tissue or organ that can self-renew and are responsible for the development and maintenance of the tissue's architecture—have the capacity to initiate a tumor, if and when they acquire sufficient number of mutations.
Premise 7. Stem cells make up a small proportion of the total number of cells in a tissue; and each tissue type has a different number and division pattern.
Premise 8: The more often a cell turns over, the more mutations they are likely to acquire (by premises 1–2).
Conclusion 2. There should be a strong, quantitative correlation between the lifetime number of divisions among a particular class of cells within each organ (stem cells) and the lifetime risk of cancer arising in that organ. (from 1 to 8).
Premise 9. There is a 65% correlation between rates of stem cell division in a given tissue type and lifetime cancer risk in that tissue type.
Conclusion 3. The differences in lifetimes risk across tissue types are caused primarily by "luck"—stochastic acquisition of mutations in stem cells.

Second version (drawing upon Schupbach 2016):

> … the hypothesis that offers the most powerful potential explanation of some proposition will be the one that makes that proposition the most likely. In Bayesian terms, the hypothesis judged to provide the best explanation will have the greatest corresponding likelihood of any explanatory hypothesis considered. This result clarifies the nature of the reason that favors the most explanatory hypothesis over those that are explanatorily inferior. A hypothesis's likelihood … is positively related to its overall probability in light of the evidence…

So, in this case the argument above may be separated into two parts: a rehearsal of the relevant background knowledge, and an assessment of the relative likelihood. The background knowledge in this case is: rates of turnover of stem cells in all tissue vary across tissues/organs, rates of mutation are constant, and cancer is (in large part) a product of the acquisition of a series of mutations over our lifetimes, there's a

65% correlation between rate of turnover of stem cells and average cancer incidence in any given tissue.

So, here's the observation (e): there are orders of magnitude of difference in cancer incidence across different tissues and organs.

What explains this?

- H1: rate of turnover of stem cells in varies across tissues/organs.
- H2, H3, etc.: exogenous factors have differential effect across tissues/organs
- Background b: rates of mutation are constant, and cancer is (in large part) a product of the acquisition of a series of mutations over our lifetimes; also perhaps: higher turnover rates increase the chances of getting the cancer-causing series of mutations (is that right?), etc.

Argument, v2

- E: There are orders of magnitude difference between cancer incidence across different tissues/organs (e.g., bone cancer and brain cancer are very rare, whereas skin cancer is relatively common, etc.)
- Out of the available potential explanations of E, H1 is the best in the sense of being the most powerful AND being the best confirmed by background evidence; formally, H1 has a much higher degree of explanatory power over E than any of the other H's, which coincides with the probabilistic claim that $Pr(E|H1\&b) >> Pr(E|Hi\&b)$ for all $i \neq 1$. AND H1 is more probable in light of b than any of the other H's: $Pr(H1|b) > Pr(Hi|b)$ for all $i \neq 1$.
- Therefore, H1.

References

Aktipis CA, Boddy AM, Jansen G, Hibner U, Hochberg ME, Maley CC, Wilkinson GS (2015) Cancer across the tree of life: cooperation and cheating in multicellularity. Philos Trans Royal Soc B: Biol Sci 370(1673):20140219

Batlle E, Clevers H (2017) Cancer stem cells revisited. Nat Med 23(10):1124

Batterman RW (2011) The tyranny of scales. http://philsci-archive.pitt.edu/8678/1/Bridging.pdf

Bechtel W (2018) The importance of constraints and control in biological mechanisms: insights from cancer research. PhilosSci 85(4):573–593

Bertolaso M (2016) Philosophy of cancer. Springer, Dordrecht

Bozic I, Antal T, Ohtsuki H et al (2010) Accumulation of driver and passenger mutations during tumor progression. ProcNatlAcadSci USA 107(43):18545–18550

Brigandt I (2013) Integration in biology: philosophical perspectives on the dynamics of interdisciplinarity. Stud HistPhilosBiol Biomed Sci 44:461–465

Broadbent A (2011a) Inferring causation in epidemiology: mechanisms, black boxes, and contrasts. In: Illari PMK, Williams J (eds) Causality in the sciences. Oxford University Press, Oxford, pp 45–69

Broadbent A (2011b) Epidemiological evidence in proof of specific causation. Leg Theory 17(4):237–278

Clevers H (2018) Bordeaux workshop in philosophy of cancer. Pers Commun

Collins AR (2005) Antioxidant intervention as a route to cancer prevention. Euro Jour of Can 41(13):1923–1930

Davy-Smith G, Relton CL, Brennan P (2016) Chance, choice and cause in cancer aetiology: individual and population perspectives. Intl J Epidemiol 45(3):605–613
Doll R, Hill AB (1950) Smoking and carcinoma of the lung. BMJ 2(4682):739
Doll R, Hill AB (1954) The mortality of doctors in relation to their smoking habits. BMJ 1(4877):1451
Duesberg P, Rasnick D (2000) Aneuploidy, the somatic mutation that makes cancer a species of its own. Cell Motil Cytoskelet 47(2):81–107
Fagan M (2013) Philosophy of stem cell biology: knowledge in flesh and blood. Palgrave Macmillan, New York
Franklin-Hall LR (2016) High-level explanation and the interventionist's "variables problem." Br J PhilosSci 67(2):553–577
Fromme JC, Verdine GL (2004) Base excision repair. Adv Protein Chem 69:1–41
Gold B (2017) Somatic mutations in cancer: Stochastic versus predictable. Mutat Res Genetic Toxicol Environ Mutagen 187:37–46
Gröbner SN, Worst BC, Weischenfeldt J et al (2018) The landscape of genomic alterations across childhood cancers. Nature 555(7696):321–327
Haldane JBS (1964) A defense of beanbag genetics. PerspectBiol Med 7(3):343–360
Hanahan D, Weinberg RA (2011) Hallmarks of cancer: the next generation. Cell 144(5):646–674
Hill AB (1965) The environment and disease: association or causation? Proc R Soc Med 58:295–300
Hochberg ME, Noble RJ (2017) A framework for how environment contributes to cancer risk. EcolLett 20(2):117–134
Hoefer C (2016) Causal determinism. In: Zalta EN (ed) The stanford encyclopedia of philosophy. https://plato.stanford.edu/archives/spr2016/entries/determinism-causal/. Accessed Spring 2020
Hu C, Hart SN, Bamlet WR et al (2016) Prevalence of pathogenic mutations in cancer predisposition genes among pancreatic cancer patients. Cancer EpidemiolPrevBiomark 25(1):207–211
Ismael JT (2016) How physics makes us free. Oxford University Press, Oxford
Ismael JT (2017) An Empiricist's guide to objective modality. In: Slater MH, Yudell Z (eds) Metaphysics and the philosophy of science: new essays. Oxford University Press, Oxford, pp 109–125
Kimsey IJ, Petzold K, Sathyamoorthy et al (2015) Visualizing transient Watson–Crick-like mispairs in DNA and RNA duplexes. Nature 519(7543):315
Kimura M (1968) Evolutionary rate at the molecular level. Nature 217(5129):624–626
Kunkel TA (2009) Evolving views of DNA replication (in) fidelity. Cold Spring HarbSymp Quant Biol 74:91–101
Laplane L (2015) Cancer stem cells. Harvard University Press, Cambridge
Laplane L, Solary E (2019) Philosophy of biology: towards a classification of stem cells. Elife 8:e46563
Malaterre C (2007) Organicism and reductionism in cancer research: towards a systemic approach. Int Stud PhilosSci 21(1):57–73
Marcum JA (2009) Cancer: complexity, causation, and systems biology. MedicinaStoria 9(17–18):267–287
Martincorena I, Campbell PJ (2015) Somatic mutation in cancer and normal cells. Science 349(6255):1483–1489
Mayr E (1963) Animal species and evolution. Harvard University Press, Cambridge
Merlin F (2016) Weak randomness at the origin of biological variation: the case of genetic mutations. In: Pence C, Ramsey G (eds) Chance in evolution. University of Chicago Press, Chicago, pp 176–195
Nelson CM, Bissell MJ (2006) Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer. Annu Rev Cell Dev Biol 22:287–309
Noble R, Kaltz O, Nunney L, Hochberg ME (2016) Overestimating the role of environment in cancers. Cancer Prev Res 9(10):773–776
Nowak M, Waclaw B (2017) Genes, environment, and bad luck: explaining cancer risk in the statistical sense. Science 355(6331):1266–1267
Nunney L (2018) Size matters: height, cell number and a person's risk of cancer. Proc R Soc B 285(1889):20181743
Nunney L, Muir B (2015) Peto's paradox and the hallmarks of cancer: constructing an evolutionary framework for understanding the incidence of cancer. Philos Trans R Soc B BiolSci 370(1673):20150161
O'Rourke M, Crowley S, Gonnerman C (2016) On the nature of cross-disciplinary integration: a philosophical framework. Stud HistPhilosSci Part C Stud HistPhilosBiol Biomed Sci 56:62–70
Plomin R, Daniels D (1987) Why are children in the same family so different from each other? Behav Brain Sci 10:1–16
Plutynski A (2011) Four problems of abduction: a brief history. HOPOS: J IntSocHistPhilosSci 1(2):227–248
Plutynski A (2018) Explaining cancer: finding order in disorder. Oxford University Press, New York

Plutynski A (2020) Cancer modeling: the advantages and limitations of multiple perspectives. In: Massimi M, McCoy CD (eds) Understanding: Perspectivism Scientific challenges and methodological prospects. Taylor & Francis, New York
Reiss J (2015) A pragmatist theory of evidence. PhilosSci 82(3):341–362
Roberts SA, Gordenin DA (2014) Hypermutation in human cancer genomes: footprints and mechanisms. Nat Rev Cancer 14(12):786–800
Rozhok AI, Wahl GM, DeGregori J (2015) A critical examination of the "bad luck" explanation of cancer risk. Cancer Prev Res 8(9):762–764
Salk JJ, Fox EJ et al (2010) Mutational heterogeneity in human cancers: origin and consequences. Annu Rev Pathol 5:51–75
Schupbach JN (2016) Inference to the best explanation, cleaned up and made respectable. In: McCain K, Poston T (eds) Best explanations: new essays on inference to the best explanation. Oxford University Press, Oxford, pp 39–61
Shapiro LA, Sober E (2012) Against proportionality. Analysis 72(1):89–93
Sober E (2010) Evolutionary theory and the reality of macro-probabilities. In: Eells E, Fetzer JH (eds) The place of probability in science. Springer, Dordrecht, pp 133–161
Sornette D, Favre M (2015) Debunking mathematically the logical fallacy that cancer risk is just "bad luck." EPJ Nonlinear Biomed Phys 3(1):10
Sonnenschein C, Soto AM (2008) Theories of carcinogenesis: an emerging perspective. Semin Cancer Biol 18(5):372–377
Soto AM, Sonnenschein C (2005) Emergentism as a default: cancer as a problem of tissue organization. J Biosci 30(1):103–118
Soto AM, Sonnenschein C (2011) The tissue organization field theory of cancer: a testable replacement for the somatic mutation theory. BioEssays 33(5):332–340
Suárez M (2020) Philosophy of probability and statistical modelling: Cambridge elements in the philosophy of science. Cambridge University Press, Cambridge
Suárez M (2017) Propensities, probabilities, and experimental statistics. In: Massimi M, Romeijn J, Schurz G (eds) EPSA15: selected papers: European studies in philosophy of science, vol 5, pp 335–45
Tomasetti C, Vogelstein B, Parmigiani G (2013) Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation. ProcNatlAcadSci 110(6):1999–2004
Tomasetti C, Vogelstein B (2015) Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science 347:78–81
Tomasetti C, Li L, Vogelstein B (2017) Stem cell divisions, somatic mutations, cancer etiology, and cancer prevention. Science 355:1330–1334
Various authors. 2015 Letters on cancer risk. *Science.* 347(6223), 728–729. (Ashford, Nicholas A, Bauman, P, Brown, HS, et al Cancer risk: role of environment; Potter, JD, Prentice, RL Cancer risk: tumors excluded; Wild, C, Brennan, P, Plummer, M, et al Cancer risk: role of chance overstated; Song, M, Giovannucci, EL Cancer risk: many factors contribute; Gotay, C, Dummer, T, Spinelli, J Cancer risk: prevention is crucial; O'Callaghan, M Cancer risk: accuracy of literature.
Weinberg CR, Zaykin D (2015) Is bad luck the main cause of cancer? JNCI: J Natl Cancer Inst 107(7)
Weslake B (2013) Proportionality, contrast and explanation. Australas J Philos 91(4):785–797
Woodward J (2010) Causation in biology: stability, specificity, and the choice of levels of explanation. BiolPhilos 25:287–318
Woodward J (2018) Explanatory autonomy: the role of proportionality, stability, and conditional irrelevance. Synthese 1–29
Woodward J (forthcoming) Causation with a human face: normative theory and descriptive psychology
Wu S, Powers W, Zhu Y, Hannun (2016) Substantial contributions of extrinsic risk factors to cancer development. Nature 529:43